

SISTEMA PARA IDENTIFICAÇÃO E MONITORAMENTO DE ESTUDANTES EM RISCO DE EVASÃO

Baldoino Sonildo da Nóbrega¹
Joab da Silva Maia²
Marcelo A. Silva Filho³
Edivan Enéas A. Júnior⁴

RESUMO

A evasão escolar tem causado prejuízos financeiros, sociais e culturais. Em relação ao ensino superior, no Brasil, foi identificado na última década que 59% dos alunos desistiram do seu curso de ingresso. As altas taxas de abandono no ensino superior retardam o desenvolvimento do país, o qual necessita de pessoas qualificadas para evoluir em educação, pesquisa, inovação e tecnologia. Diante disto, as instituições de ensino superior necessitam de mecanismos que possam acompanhar o risco de evasão dos seus alunos, para em seguida traçar estratégias de combate a esse fenômeno. Dessa forma, o objetivo desse trabalho foi desenvolver um sistema para detecção e monitoramento de alunos em risco de evasão. A pesquisa foi realizada em parceria com uma instituição de ensino pública, a partir dos dados acadêmicos e socioeconômicos dos estudantes de graduação. Através de métodos estatísticos, foi realizada a seleção e classificação das variáveis que poderiam ser consideradas relevantes para o estudo da evasão e em seguida, foram analisados os resultados dos melhores modelos de aprendizado de máquina para esses dados. Dos vários modelos avaliados, o *XGBoost (Extreme Gradient Boosting)* foi o que melhor se adaptou aos dados com boas métricas, eficiência, calibrado e com recall próximo de 0,85. A pesquisa trouxe contribuições relevantes para a instituição analisada ao disponibilizar uma plataforma para acompanhar período a período os alunos com risco de abandono do curso. Também foi identificado que a maior proporção da evasão está ocorrendo nos períodos iniciais, entre estudantes com CRE muito baixo, alta reprovação por falta e com idade elevada.

Palavras-chave: Análise de dados, Evasão, Estatísticas, Ensino Superior.

INTRODUÇÃO

Nas duas últimas décadas, ocorreu um aumento significativo de ingressantes nas Instituições de Ensino Superior (IES), no Brasil, passando de 4 milhões para 8,3 milhões (BRASIL, 2018). Um dos fatores, foi o investimento por parte do Governo Federal no sentido de expandir os cursos presenciais e à distância. Paralelo ao aumento de matrículas no ensino superior, houve crescimento também no número de evasões, que de acordo com Gilioli (2016), a média de alunos evadidos nas IES, no Brasil, está entre 30% e 40%.

¹ Mestre pelo Curso de Inovação e Desenvolvimento da Faculdade Guararapes - UniFG, baldoino.nobrega@ifpb.edu.br;

² Graduando pelo Curso de Engenharia de Computação do Instituto Federal da Paraíba - IFPB, joab.maia@academico.ifpb.edu.br;

³ Graduando pelo Curso de Engenharia de Computação do Instituto Federal da Paraíba - IFPB, marcelo.filho@academico.ifpb.edu.br;

⁴ Graduando pelo Curso de Engenharia de Computação do Instituto Federal da Paraíba - IFPB, edivan.junior@academico.ifpb.edu.br;

Esses números elevados reforçam a importância de estudos sobre a evasão e suas causas (BARROSO; FALCÃO, 2004; ANDRIOLA; RIBEIRO; MOURA, 2005; GAIOSO, 2005; MIRANDA, 2006; SANTOS; LOPES, 2011; SOUZA, DA SILVA; GESSINGER, 2012). Essa problemática tem-se destacado, uma vez que está cada vez mais causando prejuízos financeiros, sociais e culturais (RAFAEL *et al.*, 2015). É importante destacar, que o crescimento do fenômeno da evasão prejudica o desenvolvimento do país, visto que se necessita cada vez mais de mão de obra qualificada para atingir resultados melhores em educação, ciência e tecnologia.

A literatura identifica que o problema da evasão nos cursos superiores é multifatorial (HOED, 2016). Portanto, esse fenômeno não é tão simples de combater, já que as variáveis são diversas e difíceis de controlar. De acordo com Hoed (2016) e Tinto (1975), diferentes questões podem influenciar a saída precoce do aluno: questões socioeconômicas, vocacionais, institucionais, pessoais, de saúde, familiares, políticas governamentais ligadas à educação, formas de ingresso e tipo de instituição (se privada ou pública). Todos esses fatores podem contribuir, em algum momento, para a saída do aluno. Dessa forma, cada instituição deve identificar quais desses motivos influenciam de modo significativo a evasão para se antecipar e tentar reverter a situação, quando for possível. Esses fatores precisam ser avaliados e monitorados, para que as instituições de ensino possam prever uma possível saída do aluno e, conseqüentemente tomar decisões que favoreçam a sua permanência até a conclusão do curso.

Identificar e acompanhar os alunos com risco de evasão tem se tornado um desafio para as IES, seja pela dificuldade de definir as variáveis que influenciam, ou pela falta de um mecanismo que favoreça esse acompanhamento contínuo do aluno. Uma forma para auxiliar as IES nessas perspectivas é criar um sistema de monitoramento e detecção de alunos em risco de evasão, a partir de suas bases de dados junto ao sistema de controle acadêmico. Esses dados sendo tratados e analisados, à medida que o aluno evolui no curso, podem trazer informações relevantes e possibilitar a IES se antecipar e aplicar estratégias para tentar reverter a desistência por parte do estudante.

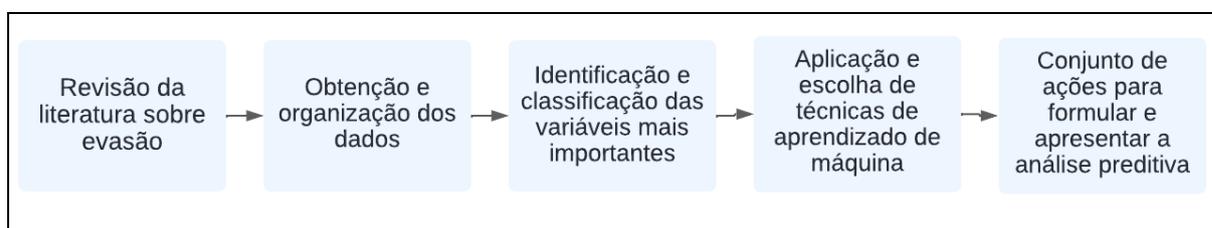
Alguns estudos têm utilizado modelos de aprendizado de máquina e obtido resultados positivos na criação de mecanismos de predição de risco de evasão (AMORIM; BARONE; MANSUR, 2008; MELO, 2016; DIGIAMPIETRI; NAKANO; LAURETTO, 2016; SCHREIBER *et al.* 2017; ARAUJO; SOUZA; SANTANA, 2018; BELTRAN; XAVIER; BARRETO; NETO, 2019). Contudo, alguns têm ficado apenas no treinamento e validação do modelo, sem necessariamente fornecer uma plataforma completa e acessível de acompanhamento para que os gestores tomem as decisões, a partir dos resultados gerados.

Diante de toda a problemática, o objetivo desse trabalho foi desenvolver um sistema de detecção e monitoramento de alunos em risco de evasão. O Sistema é composto de algoritmos de aprendizado de máquina, análise estatística e organização das informações de tal modo que demonstrem a evolução do risco de evasão e possibilite uma tomada de decisão com fundamentos nesses resultados.

METODOLOGIA

O presente trabalho assume uma abordagem quantitativa e utiliza técnicas de mineração de dados para atingir o objetivo. Foram utilizados dados acadêmicos de estudantes dos cursos de Licenciatura em Matemática, Licenciatura em Física, Construção de Edifícios, Telemática e Engenharia de Computação. Os dados foram disponibilizados pela instituição pública parceira desta pesquisa. A amostra foi composta de aproximadamente 1700 alunos que ingressaram entre os anos de 2017 e 2021. A metodologia foi dividida em 5 etapas (Figura 1).

Figura 1: Etapas da metodologia



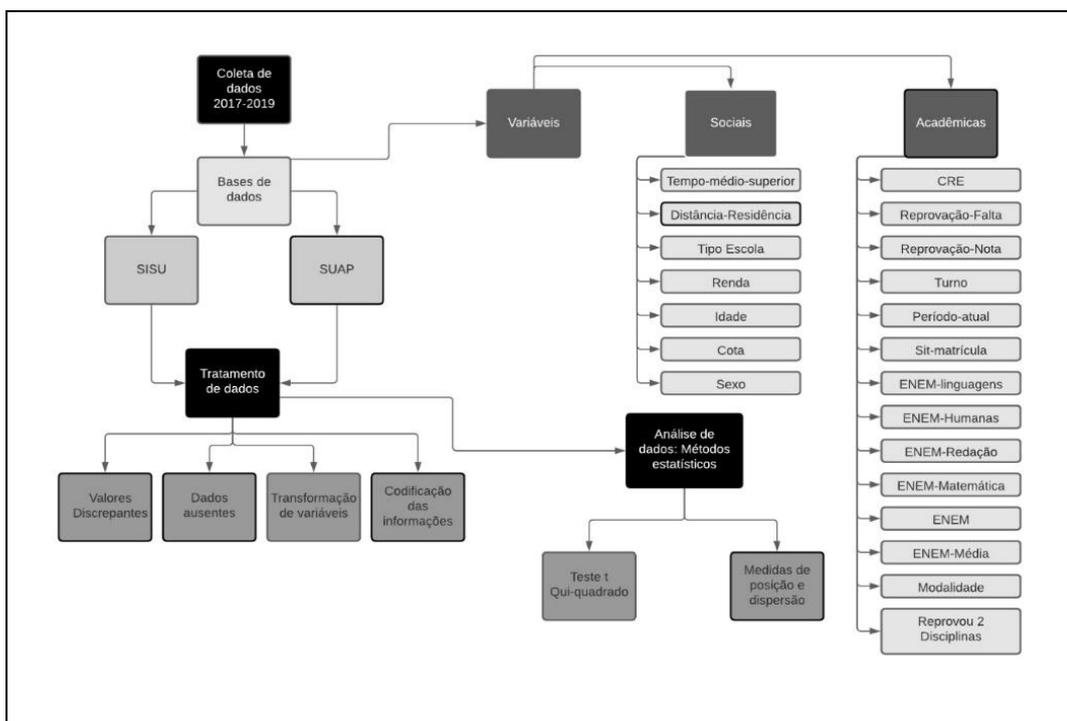
Fonte: autores, 2022.

Na primeira etapa foi realizado uma revisão da literatura para avaliar quais variáveis eram relevantes para o estudo da evasão e quais técnicas de aprendizado de máquina estão sendo utilizadas nesse tipo de problema. A segunda etapa consistiu na construção de um banco de dados estruturado a partir da junção de dados secundários disponibilizados no sistema de controle acadêmico da instituição (coeficiente de rendimento escolar-CRE, reprovações, situação no curso, cota, notas do ENEM, renda, idade, turno, entre outras) (Figura 2). Importante destacar que os alunos tiveram suas identificações preservadas, retirando da base de dados qualquer tipo de informação pessoal que identificasse o aluno.

Para o processo de tratamento de dados e análise dos modelos foi utilizado a linguagem de programação python com o apoio da *scikit-learn*, que é uma biblioteca de aprendizado de máquina de código aberto que auxilia na análise preditiva. A terceira etapa da pesquisa utilizou métodos estatísticos para a identificação e classificação de quais variáveis organizadas na

segunda etapa, podem ser consideradas relevantes para o estudo da evasão. A quarta etapa tratou da modelagem, foram realizadas a identificação e avaliação dos modelos de aprendizado de máquina mais adequados para a situação. As métricas consideradas para avaliação dos modelos foram: acurácia, precisão, *recall*, coeficiente *kappa*. É importante informar que esta pesquisa tratou a predição como um problema de classificação binária, onde os modelos classificam o objeto de estudo como evasão ou não evasão.

Figura 2: Processo de coleta e organização das variáveis



Fonte: autores, 2022

A quinta e última etapa consistiu no conjunto de ações para formular a ferramenta de detecção e monitoramento do risco de evasão, visando disponibilizar as informações de modo acessível e interpretável. Inicialmente, foi criada uma aplicação *web* que possibilitou integrar o modelo escolhido na quarta etapa e a base de dados. Essa aplicação foi restrita com necessidade de *login* e senha para garantir o acesso apenas a pessoas autorizadas, visando o sigilo das informações.

RESULTADOS E DISCUSSÃO

Análise da associação das variáveis

Na análise estatística observou-se que existia uma associação de algumas variáveis com a evasão: idade de ingresso, tempo entre ensino médio e inclusão no ensino superior, CRE, reprovação por falta, quantidade de períodos cursados e a renda com destaque apenas para categoria “não declarada”. Estas apresentaram significância estatística com $p\text{-valor} < 0,05$.

Ao avaliar as variáveis associadas a evasão, identificou-se que a média de idade dos matriculados (23 anos) é inferior aos evadidos (26 anos). O tempo entre ensino médio e a inclusão no ensino superior também é mais elevado para os alunos evadidos. Quando avaliado o desempenho acadêmico pelo CRE, verificou-se que havia uma discrepância entre evadidos (média de 20 pontos) e matriculados (média de 60 pontos). Foi observado que 75% dos evadidos tinham CRE inferior a 30 pontos. A medida de desempenho acadêmico é apresentada como preditora da evasão por alguns autores (FRITSCH; DA ROCHA; VITELLI, 2015, TINTO, 1975), portanto acompanhar essa variável é importante para verificar indícios de abandono do curso.

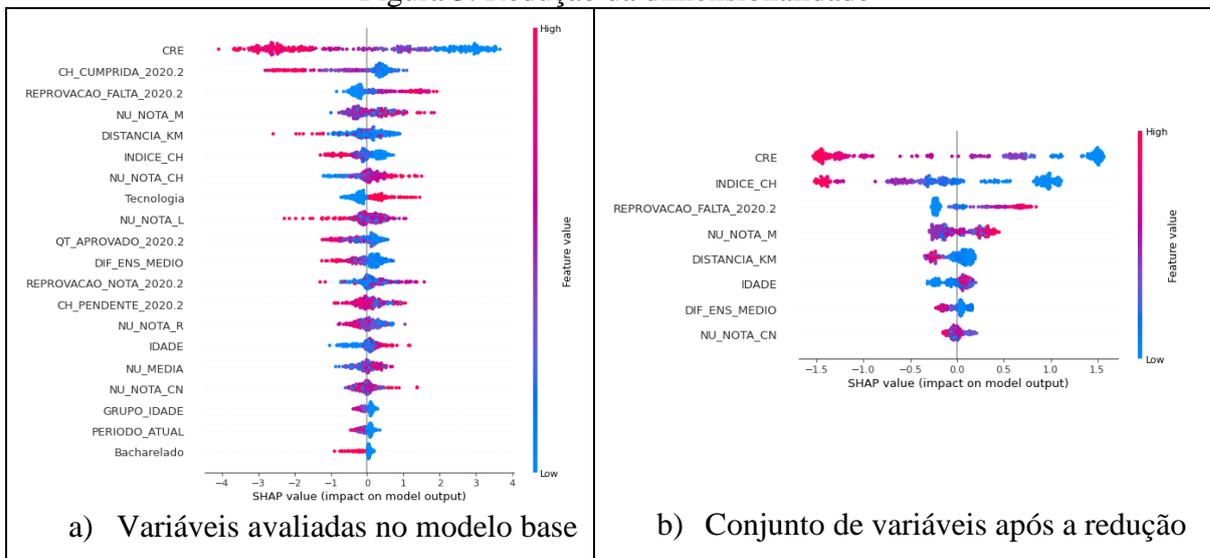
A média de reprovação entre os indivíduos evadidos (média de 3,5 disciplinas) foi superior aos matriculados (média de uma disciplina). Essas reprovações ocorreram principalmente em períodos iniciais. Verificou-se que 70% dos estudantes não conseguem cursar dois períodos, o que demonstra uma evasão intensa no início do curso. Esses resultados convergem com estudos realizados por Gilioli (2016) e Brasil (2019).

Seleção dos atributos

Para avaliar os modelos de aprendizado de máquina, primeiro buscou-se reduzir o número de variáveis, visto que uma quantidade alta dessas, resulta em um modelo complexo. Desse modo, foram realizados experimentos de seleção de atributos para visualizar, interpretar e escolher quais impactam de forma relevante na decisão do modelo. Um dos métodos utilizados foi o *shapley additive explanations*.

A partir de um modelo base, observou-se como os atributos se comportavam (Figura 3 a). Após testar diversas combinações de atributos, os resultados mostraram que o melhor conjunto era composto por oito variáveis (Figura 3 b). Com essas, o modelo continuou performando com boas métricas e menos complexo na interpretação.

Figura 3: Redução da dimensionalidade



Fonte: autores, 2022

Ao avaliar os resultados obtidos na análise de associação das variáveis e observar quais atributos impactam o modelo de aprendizado de máquina, foram selecionados: CRE, índice de carga horária cursado, reprovação por falta, nota do ENEM em matemática, a distância percorrida até a IES, idade de ingresso, diferença entre o ensino médio e inclusão no ensino superior e a nota do ENEM em ciências da natureza (Figura 3 b).

Avaliação dos Modelos

Na parte de avaliação dos modelos foi utilizada a validação cruzada, que é uma forma de treinar e testar os modelos em vários grupos, podendo ser definidos de forma aleatória, o que favoreceu a redução de um possível viés. Para determinar qual modelo de aprendizado de máquina se adequava melhor aos dados, foram realizados vários experimentos, sendo apresentados apenas os modelos que possuíam as melhores métricas (Quadro 1).

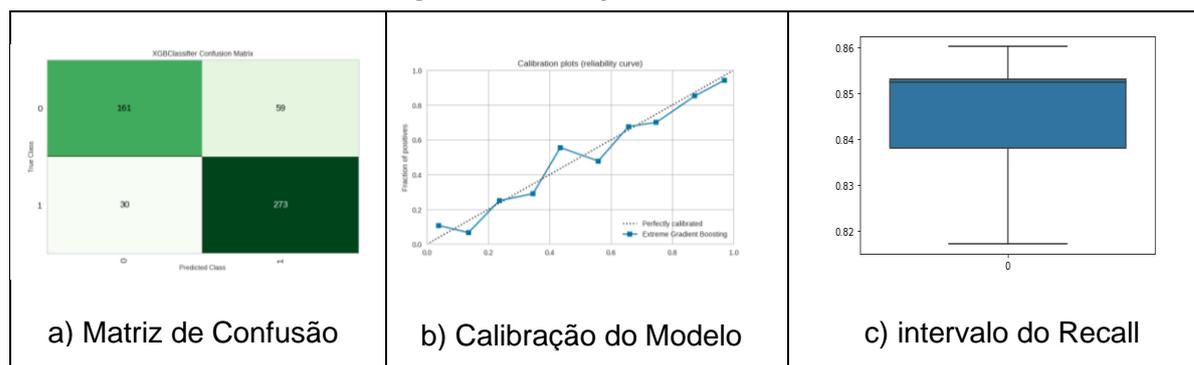
Quadro 1: Resultados das métricas para avaliação dos modelos

Modelo	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
<i>Extreme Gradient Boosting</i>	0.8302	0.9087	0.8530	0.8356	0.8434	0.6579	0.8302
<i>Logistic Regression</i>	0.8384	0.9124	0.8469	0.8539	0.8492	0.6751	0.6771
<i>CatBoost Classifier</i>	0.8384	0.9124	0.8469	0.8539	0.8492	0.6751	0.6771
<i>Random Forest Classifier</i>	0.8343	0.9129	0.8422	0.8491	0.8446	0.6670	0.6690
<i>Linear Discriminant Analysis</i>	0.8318	0.9127	0.8346	0.8523	0.8422	0.6622	0.6644

Fonte: autores, 2022

Dentre todas as métricas considerou-se o *recall* como medida de decisão. Dessa forma o modelo considerado foi o *Extreme Gradient Boosting (XGBoost)* (Quadro 1). A escolha foi feita após a análise da validação cruzada para 10 partições, no conjunto de dados. No grupo de teste, o *XGBoost* se manteve bem calibrado (Figura 5 b) e com boas métricas, chegando a atingir *recall* próximo de 0.85 (Figura 5 c).

Figura 5: Avaliação dos modelos



Fonte: autores, 2022

O modelo *XGBoost* é um dos principais algoritmos usados atualmente em competições de *kaggle* e consegue lidar com robustez em diversos tipos de dados. A motivação pela escolha desse modelo foi tanto pela adaptação aos dados quanto pela eficiência e calibração (Figura 5).

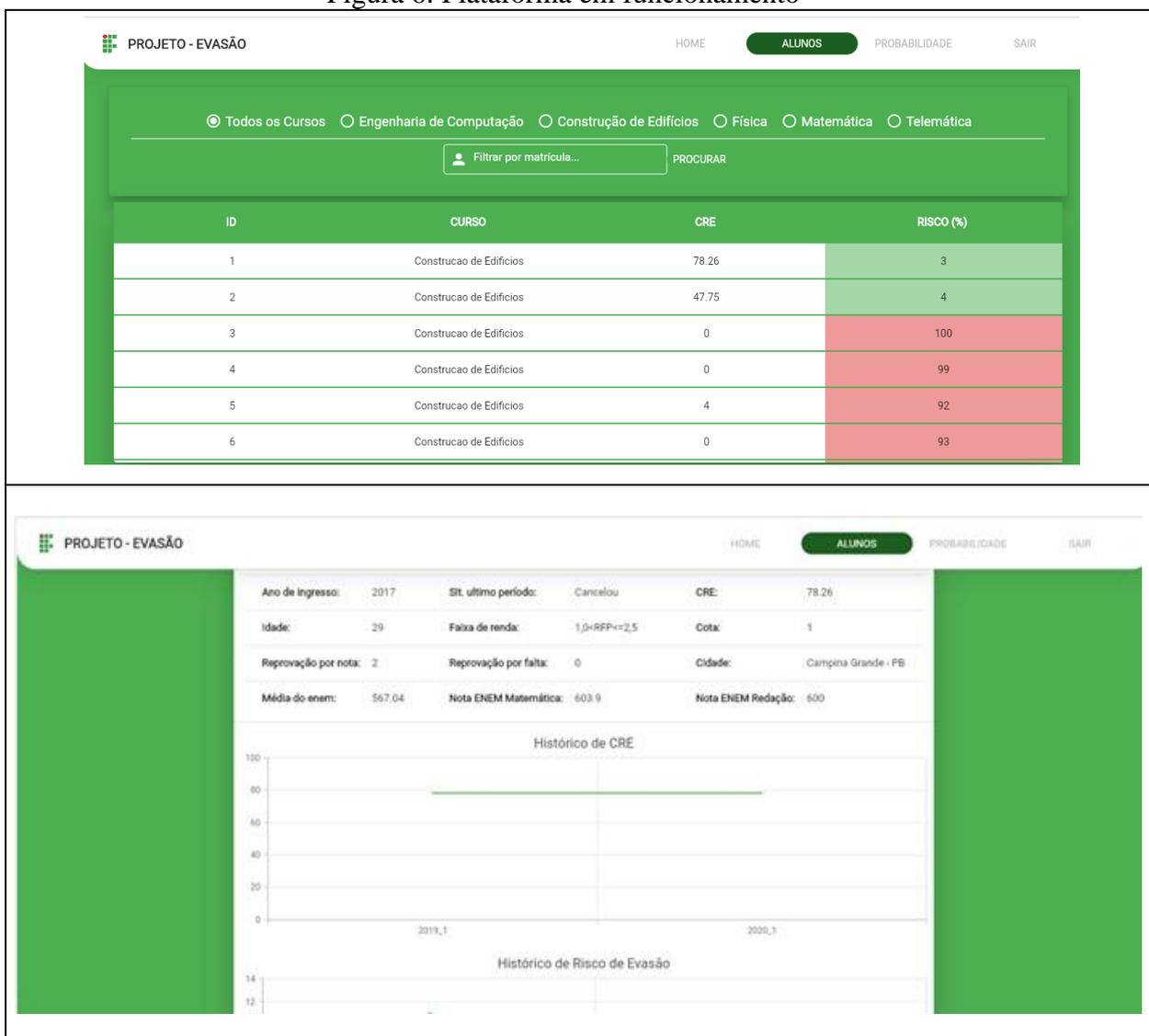
Plataforma de Integração dos dados e análise de risco

A plataforma desenvolvida utiliza dados acadêmicos e socioeconômicos para tomar decisões sobre alunos com risco de evasão. Um modelo de aprendizado de máquina treinado está integrado para avaliar e tomar as decisões sobre a chance de evasão de um aluno. Através de gráficos, observa-se a evolução do aluno no curso e se o risco está aumentando ou diminuindo com o passar do tempo.

Algumas funcionalidades da plataforma são (Figura 6): 1. Listar os alunos com risco de evasão; 2. Monitorar o risco período por período; 3. Verificar as principais variáveis, por exemplo, o CRE durante o curso, risco de evasão, renda, notas do ENEM, ou seja, possibilita traçar o perfil do aluno; 4. Painel de dados na tela inicial para acompanhamento e visualização das principais características do público; 5. Listagem com filtros por curso ou por características de evasão; 6. Possibilidade de entrar com as características do aluno e medir o risco de evasão

automaticamente; 7. Acessar a plataforma através de *login* e senha para manter a segurança dos dados. Toda essa estrutura está organizada e integrada em um servidor de aplicação *web*.

Figura 6: Plataforma em funcionamento



Fonte: autores, 2022

As diversas funcionalidades desenvolvidas utilizaram o serviço de banco de dados *PostgreSQL*, o *framework NestJS/NodeJs* para o *backend* e *flutter* para o *frontend*, além da *API REST*. Com essa estrutura conectada, foi possível ter a plataforma *web* em funcionamento, o que possibilitou a integração do modelo de decisão e visualização dos dados no painel de apresentação (Figura 6).

É importante ressaltar que, a plataforma criada servirá como um alerta e possibilitará o acompanhamento personalizado para alunos com risco de evasão. Os profissionais que atuam na gestão do ensino poderão, a partir dos resultados, traçar estratégias para minizar o abandono

do curso e melhorar a tomada de decisão. Os dados obtidos na plataforma poderão funcionar como um ponto de partida para uma análise detalhada sobre a evasão na instituição de ensino, dessa forma, essas possibilidades demonstram o potencial do sistema desenvolvido.

CONSIDERAÇÕES FINAIS

A pesquisa teve por objetivo desenvolver um sistema para detecção e monitoramento de alunos em risco de evasão. A partir da análise descritiva dos resultados, observou que algumas variáveis podem servir como ponto de partida para identificar perfis de alunos com risco de evasão do curso. As variáveis com algum tipo de associação com a evasão foram: idade de ingresso, tempo entre Ensino Médio e inclusão no Ensino Superior, CRE, reprovação por falta, quantidade de períodos cursados e a renda com destaque apenas para categoria “não declarada”.

Após a análise das variáveis testou-se vários algoritmos de aprendizagem de máquina, com uma abordagem supervisionada, para classificação quanto ao risco de evasão escolar. Os resultados mostraram ao menos que cinco modelos construídos são efetivos na predição quanto a evasão, para esse tipo de conjunto de dados. O modelo escolhido e implementado para realizar as predições foi o *XGBoost*, pois manteve-se com boas métricas, calibrado e consistente nos dados de teste.

A partir do sistema implementado e integrado no servidor foi possível conectar o banco de dados, o modelo de decisão e o painel de informações que compila todas as características do aluno e gera relatórios capaz de auxiliar na tomada de decisão, quanto ao risco de evasão para cada aluno. Uma das contribuições do sistema será oferecer aos gestores da instituição de ensino a possibilidade do monitoramento dos alunos com alto risco de evasão e assim tomar medidas antecipadas com vista a reduzir esse fenômeno. Como trabalhos futuros, seria importante montar uma estratégia para que o sistema faça indicações de matérias pedagógicas para melhorar o aprendizado dos alunos que estão em risco de evasão.

AGRADECIMENTOS

Ao Instituto Federal de Educação Ciência e Tecnologia da Paraíba, pelo apoio integral para o desenvolvimento deste estudo, através do Programa de Apoio a Projetos de Pesquisa, Inovação e Desenvolvimento Tecnológico e Social (Interconecta). Edital nº 02/2021.

REFERÊNCIAS

AMORIM, M. J. V.; BARONE, D.; MANSUR, A. U. Técnicas de aprendizado de máquina aplicadas na previsão de evasão acadêmica. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2008. p. 666-674.

ANDRIOLA, B. A.; RIBEIRO, E. S.; MOURA, C. P. **Evasão discente nos cursos de graduação da Universidade Federal do Ceará (UFC): busca das suas causas**. In: ANDRIOLA, W. B. (Org.). Avaliação: múltiplos olhares em educação. Fortaleza: Ed. da Universidade Federal do Ceará, 2005

ARAÚJO J. N. R.; SOUZA, W. S.; SANTANA, S. A. Proposta de Ferramenta para Predição de Evasão no Ensino Superior. **e-RAC**, v. 8, n. 1, 2018.

BARROSO, M. F.; FALCÃO, E. B. M. **Evasão universitária: o caso do Instituto de Física da UFRJ**. IX Encontro Nacional de Pesquisa em Ensino de Física, v. 9, p. 1-14, 2004.

BELTRAN C. A. R.; XAVIER, J. C.J; BARRETO, C. A. S.; NETO, C. A. O. Plataforma de aprendizado de máquina para detecção e monitoramento de alunos com risco de evasão. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. 2019. p. 1591.

BRASIL. Ministério da Educação. Instituto Nacional De Estudos E Pesquisas Educacionais Anísio Teixeira – INEP. **Censo da Educação Superior 2017**. Brasília, 2018. Disponível em:<<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior>>. Acesso em: 15 abril 2022.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. **Censo da Educação Superior 2019**. Brasília, 2019. Disponível em:<https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/resumo_tecnico_censo_da_educacao_superior_2019.pdf>. Acesso em: 21 maio 2022.

DIGIAMPIETRI, L. A.; NAKANO, F.; SOUZA LAURETTO, M. Mineração de dados para identificação de alunos com alto risco de evasão: Um estudo de caso. **Revista de Graduação USP**, v. 1, n. 1, p. 17-23, 2016.

FRITSCH, Rosângela; DA ROCHA, Cleonice Silveira; VITELLI, Ricardo Ferreira. A evasão nos cursos de graduação em uma instituição de ensino superior privada. **Revista Educação em Questão**, v. 52, n. 38, p. 81-108, 2015.

GAIOSO, N. P. L. **Evasão discente na educação superior: a perspectiva dos dirigentes e dos alunos**. Dissertação (Mestrado) - Universidade Católica de Brasília - Brasília, Brasil, 2005.

GILIOLI, R. S. P. Evasão em instituições federais de ensino superior no Brasil: Expansão da rede, SISU e desafios. Brasília: **Consultoria Legislativa: Câmara dos Deputados**, 2016.

HOED, R. M. **Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de Computação**. Orientador: Marcelo Ladeira. 2016. 188 f. Dissertação (Mestrado em Computação Aplicada) - Universidade de Brasília. Brasília. 2016.

MELO, A. S. C. **Previsão automática de evasão estudantil: um estudo de caso na UFCG. Campina Grande, UFCG**. Orientador: Leandro Balby Marinho. 2016. 63 f. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Campina Grande. Campina Grande-PB, 2016

MIRANDA, D. P. M. **Gestão da evasão nas instituições de ensino superior privado: um estudo sobre cursos de Administração no Estado do Espírito Santo.** Dissertação (Mestrado) - Universidade Federal Rural do Rio de Janeiro, Instituto de Ciências Humanas e Sociais. 104 p. 2006

RAFAEL, J. A. M. et al. Análise da Evasão em um Curso de Licenciatura em Matemática da Rede Federal de Ensino nos Seus Primeiros Cinco Anos de Implantação. In: **Revista Paranaense de Educação Matemática**, Campo Mourão, v. 4, n. 6, p. 118-135, jan./fev. 2015.

SANTOS B; LOPES, D. A. **Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica.** Avaliação: Revista da Avaliação da Educação Superior (Campinas), v. 16, n. 2, p. 355-374, 2011.

SCHREIBER, J. N. C. BESKOW I. A.; NARA E. O. B.; SILVA J. I.; MORAES, J.; NAJDZION V. M.; **Software Sdbayes: Um Auxílio Para A Predição De Evasão Discente.** In: XVII Colóquio Internacional de Gestão Universitária, 2017, Argentina. Anais Argentina, 2017.

SOUZA, C. T. DA SILVA, C.; GESSINGER, R. M. Um estudo sobre evasão no ensino superior do Brasil nos últimos dez anos. In: **Congressos CLABES**. 2012

TINTO, V. **Dropout from higher education: a theoretical synthesis of recent research.** Review of Educational Research, Washington, v. 45, n.1, p.89-125, 1975.